NONLINEAR WAVELET ESTIMATION OF TIME-VARYING AUTOREGRESSIVE PROCESSES

Rainer Dahlhaus[1], Michael H. Neumann[2] and Rainer von Sachs[3]

ABSTRACT. We consider nonparametric estimation of the coefficients $a_i(\cdot)$, $i = 1, \ldots, p$, of a time-varying autoregressive process. Choosing an orthonormal wavelet basis representation of the functions $a_i(\cdot)$, the empirical wavelet coefficients are derived from the time series data as the solution of a least squares minimization problem. In order to allow the $a_i(\cdot)$ to be functions of inhomogeneous regularity, we apply nonlinear thresholding to the empirical coefficients and obtain locally smoothed estimates of the $a_i(\cdot)$. We show that the resulting estimators attain the usual minimax L_2-rates up to a logarithmic factor, simultaneously in a large scale of Besov classes.

1991 Mathematics Subject Classification. Primary 62M10; secondary 62F10
Key words and phrases. Nonstationary processes, time series, wavelet estimators, time-varying autoregression, nonlinear thresholding

[1]Institut für Angewandte Mathematik, Universität Heidelberg
[2]Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin
[3]Fachbereich Mathematik, Universität Kaiserslautern

1. Introduction

Stationary models have always been the main focus of interest in the theoretical treatment of time series analysis. For several reasons autoregressive models form a very important class of stationary models: They can be used for modeling a wide variety of situations (for example data which show a periodic behavior), there exist several efficient estimates which can be calculated via simple algorithms (Levinson-Durbin algorithm, Burg-algorithm), the asymptotic properties including the properties of model selection criteria are well understood.

Frequently, people have tried to use autoregressive models also for modeling data that show a nonstationary behavior, mainly by fitting AR-models on small segments. This method is for example often used in signal analysis for coding a signal (linear predictive coding) or for modeling data in speech analysis. The underlying assumption then is that the data are coming from an autoregressive process with time varying coefficients.

Suppose we have some observations $\{X_1, \ldots, X_T\}$ from a zero mean, autoregressive process with time varying coefficients $a_1(\cdot), \ldots, a_p(\cdot)$. To get a tractable frame for our asymptotic analysis we assume that the functions a_i are supported on the interval $[0,1]$ and connected to the underlying time series by an appropriate rescaling. This leads to the model

$$X_{t,T} + \sum_{i=1}^{p} a_i(t/T) X_{t-i,T} = \sigma(t/T)\varepsilon_t, \tag{1.1}$$

where the ε_t's are independent, identically distributed with $E\varepsilon_t = 0$ and $var(\varepsilon_t) = 1$. This time varying autoregressive model is a special locally stationary process as defined in Dahlhaus (1993). However, for the main results of this paper we only use the representation (1.1) and not the general properties of a locally stationary process.

The estimation problem now consists of estimating the parameter functions $a_i(\cdot)$. Very often these functions are estimated at a fixed time point t_0/T by fitting a stationary model in a neighborhood of t_0, e.g. by estimating $a_1(t_0/T), \ldots, a_p(t_0/T)$ with the classical Yule-Walker (or Burg-) estimate over the segment $X_{t_0-N,T}, \ldots, X_{t_0+N,T}$ where N/T is small. This method has the disadvantage that it automatically leads to a smooth estimate of $a_i(\cdot)$. Sudden changes in the $a_i(\cdot)$, as they are quite common e.g. in signal analysis, cannot be detected by this method. Moreover, the performance of this method depends on the appropriate choice of the segmentation parameter N. Instead, in this paper we develop an automatic alternative, which avoids this a priori choice and adapts to local smoothness characteristics of the $a_i(\cdot)$.

Our approach consists in a nonlinear wavelet method for the estimation of the coefficients $a_i(\cdot)$. This concept, based on orthogonal series expansions, has recently been entered in the nonparametric regression estimation problem due to Donoho and Johnstone (1992) and has been proven very useful if the class of considered

functions to be estimated exhibits a varying degree of smoothness. Some generalizations can be found in Brillinger (1994), Johnstone and Silverman (1994), Neumann and Spokoiny (1995) and Neumann and von Sachs (1995a). As usual, the unknown functions, i.e. $a_i(u)$, are expanded by orthogonal series w.r.t. a particularly chosen orthonormal basis of $L_2[0,1]$, a *wavelet* basis. Basically, the basis functions are generated by dilations and translations of the so-called scaling function ϕ and wavelet function ψ, which are both localized in spatial position (i.e. temporial, here) and frequency. These basis functions, unlike most of the "traditional" ones (Fourier, (non-local) polynomials, etc.), are able to optimally compress both functions with rather homogeneous smoothness over the whole domain (like Hölder or L_2-Sobolev) as well as members of certain inhomogeneous smoothness classes like L_p-Sobolev or Besov $B^m_{p,q}$ with $p < 2$. Note that the better compressed a signal is (i.e. being represented by a smaller number of coefficients), the better performs an estimator of the signal which is optimally tuned w.r.t. bias-variance trade-off. A strong theoretical justification for the merits of using wavelet bases in this context has been given by Donoho (1993): It was shown that wavelets provide unconditional bases for a wide variety of these inhomogeneous smoothness classes which yields that wavelet estimators can be optimal in the abovementioned sense.
To actually achieve this optimality there is need for non-linearly modifying traditional linear series estimation rules which are known to be optimal only in case of homogeneous smoothness: There the coefficients of each resolution level j are essentially of the same order of magnitude, and the loss due to a levelwise inclusion/exclusion rule, as opposed to a componentwise rule, is only small. However, under strong inhomogeneity, not only the coefficients of each fixed level might considerably differ in their order of magnitude but also have significant coefficients on higher levels to be included by a suitably chosen inclusion rule. Surprisingly enough, this is possible by simple and intuitive schemes which are based on comparing the size of the empirical (i.e. estimated) coefficents with their variability. Such nonlinear rules can dramatically outperform linear ones for the mentioned cases of sparse signals (i.e. those of inhomogeneous function classes being represented in an unconditional bases).

In this work, we apply these locally adaptive estimation procedures to the particular problem of autoregression coefficients which are functions of time. In a first step, the empirical wavelet coefficients are derived as a solution of a least squares minimization problem, before, secondly, soft or hard thresholding is applied. We show that in this situation, which is considerably more complicated than ordinary regression, our nonlinear wavelet estimator attains the usual near-optimal minimax rate of L_2-convergence, in a large scale of Besov classes.

Finally, with this adaptive estimation of the time-varying autoregression coefficients, we immediately provide a parametric estimate for the resulting time- dependent spectral density of the process given by (1.1). An alternative, fully nonparametric approach for estimating the so-called evolutionary spectrum of a general locally stationary process (as defined in Dahlhaus (1993)) has been delivered by Neumann and von Sachs (1995b), which is based on nonlinear thresholding in a two-dimensional wavelet basis.

The content of our paper is organized as follows: While in the next section we describe details of our set-up and present this main result, in Section 3 the statistical properties of the empirical coefficients are given. Section 4 deals with the proof of the main theorem. The remaining Sections 5 - 7 collect some auxiliary results, both of own interest and in this particular context used to derive the main proof (of Section 4).

2. Assumptions and the main result

Before we develop nonlinear wavelet estimators for the functions a_i, we describe the general set-up. It is well-known that the boundary-corrected Meyer wavelets (Meyer (1991)) or those developed by Cohen, Daubechies and Vial (1993) form orthonormal bases of $L_2[0,1]$. Accordingly, we can expand a_i in an orthogonal series

$$a_i \;=\; \sum_{k\in I_l^0} \alpha_{lk}^{(i)}\phi_{lk} \;+\; \sum_{j\geq l}\sum_{k\in I_j} \beta_{jk}^{(i)}\psi_{jk}, \tag{2.1}$$

where $\alpha_{lk}^{(i)} \;=\; \int a_i(u)\phi_{lk}(u)\,du$, $\beta_{jk}^{(i)} \;=\; \int a_i(u)\psi_{jk}(u)\,du$ are the usual Fourier coefficients, also called wavelet coefficients. It is known that $\#I_j = 2^j$ and $\#I_l^0 = 2^l + N$ for some integer N depending on the regularity of the wavelet basis.

Assume a degree of smoothness m_i for the function a_i. In accordance with this, we choose compactly supported wavelet functions of regularity $r > m := \max\{m_i\}$, that is

(A1) (i) ϕ and ψ are $C^r[0,1]$ and have compact support,
(ii) $\int \phi(t)\,dt = 1, \quad \int \psi(t)t^k\,dt = 0 \quad \text{for} \quad 0\leq k\leq r.$

By the usual approach, as derived in the abovementioned work on boundary-corrected wavelets, we now obtain basis functions of $L_2[0,1]$ as $\phi_{lk} = 2^{l/2}\phi(2^l x - k)$ and $\psi_{jk} = 2^{j/2}\psi(2^j x - k)$, with certain modifications of those functions that have a support beyond the interval $[0,1]$.

The first step in each wavelet analysis is the definition of empirical versions of the wavelet coefficients. Here we obtain such coefficients $\widetilde{\alpha}_{lk}^{(i)}$ and $\widetilde{\beta}_{jk}^{(i)}$ as least squares estimators corresponding to some truncated wavelet series expansion of the functions a_i; see Section 3 for a detailed description of that procedure.

To treat these coefficients in a statistically appropriate manner, we have to tune the estimator in accordance with their distribution. It turns out that this distribution actually depends on the (unknown) distribution of the $X_{t,T}$'s at the finest resolution scales, whereas we can hope to have asymptotic normality if $2^j = o(T)$. We show in Section 3 that we do not lose asymptotic efficiency of the estimator, if we truncate the series at some level $j = j(T)$ with $2^{-j(T)} = O(T^{-1/2})$. To give a definite rule, we choose the highest resolution level j^*-1 such that $2^{j^*-1} \leq T^{1/2} < 2^{j^*}$, i.e. we restrict our analysis to coefficients $\widetilde{\alpha}_{lk}^{(i)}$ ($k\in I_l^0$, $i=1,\ldots,p$) and $\widetilde{\beta}_{jk}^{(i)}$ ($j\geq l$, $2^{j+1}\leq T^{1/2}$, $k\in I_j$, $i=1,\ldots,p$). Unlike in ordinary regression it is not possible in the autocorrelation problem considered here to include coefficients from resolution scales j up to $2^j \asymp T$. This is due to the fact that the empirical coefficients cannot be reduced to sums of independent (or sufficiently weakly dependent) random variables, which results in some additional bias term.

In the present paper we propose to apply nonlinear smoothing rules to the coefficients $\widetilde{\beta}_{jk}^{(i)}$. It is well-known (cf. Donoho and Johnstone (1992)) that linear estimators can be optimal w.r.t. the optimal rate of convergence as long as the underlying smoothness of a_i is not too inhomogeneous. This situation changes considerably, if the smoothness varies strongly over the domain. Then we have the new effect that even at higher resolution scales a small number of coefficients cannot be neglected, whereas the overwhelming majority of them is much smaller than the noise level. This kind of sparsity of the signal is responsible for the need of a nonlinear estimation rule. Two commonly used rules to treat the coefficients are

1) hard thresholding
$$\delta^{(h)}(\widetilde{\beta}_{jk}^{(i)}, \lambda) \;=\; \widetilde{\beta}_{jk}^{(i)} I\left(|\widetilde{\beta}_{jk}^{(i)}| \geq \lambda\right)$$

and

2) soft thresholding
$$\delta^{(s)}(\widetilde{\beta}_{jk}^{(i)}, \lambda) \;=\; \left(|\widetilde{\beta}_{jk}^{(i)}| - \lambda\right)_+ sgn(\widetilde{\beta}_{jk}^{(i)}).$$

Before we state our main result, we introduce some more assumptions. The constant C used here and in the following is assumed to be positive, but need not be the same at each occurrence.

(A2) There exists some $\gamma \geq 0$ such that
$$|cum_n(\varepsilon_t)| \;\leq\; C^n (n!)^{1+\gamma} \qquad \text{for all} \quad n, t$$

(A3) The process $\{X_{t,T}\}$ admits an MA(∞)-representation
$$X_{t,T} \;=\; \sum_{s=0}^{\infty} \gamma_{t,T}(s) \varepsilon_{t-s}$$
with
$$\sum_{s=0}^{\infty} \sup_t \{|\gamma_{t,T}(s)|\} \;\leq\; C \qquad \text{for all} \quad T.$$

(A4) The a_i and σ are uniformly continuous with $C_1 \leq \sigma(s) \leq C_2$ on $(0,1)$ and there exists a $\delta > 0$ with
$$1 \,+\, \sum_{i=1}^{p} a_i(s) z^i \neq 0 \quad \text{for all} \;\; |z| \leq 1 + \delta \;\; \text{and all} \;\; s \in [0,1].$$

Remark 1. Note that, besides the obvious case of the normal distribution, many of the distributions that can be found in textbooks satisfy (A2) for an appropriate choice of γ. In Johnson and Kotz (1970) we can find closed forms of higher order cumulants of the exponential, gamma, inverse Gaussian and F-distribution, which show that this condition is satisfied for $\gamma = 0$. The need for a positive γ occurs in the case of heavier-tailed distribution, which could arise as the distribution of a sum of weakly dependent random variables.

(A4) implies uniform continuity of the covariances of $\{X_{t,T}\}$ (Lemma 7.1). We conjecture that the continuity in (A4) can e.g. be relaxed to piecewise continuity.

Furthermore, we conjecture that (A4) implies (A3).

In the following we derive a rate for the risk of the proposed estimator uniformly over certain smoothness classes. It is known that wavelet bases induce a norm in the space of coefficients which is equivalent to the norm in a Besov space $B^m_{p,q}$. Here $m \geq 1$ denotes the degree of smoothness and p, q $(1 \leq p, q \leq \infty)$ are shape parameters. Fix any positive constants C_{ij}, $i = 1, \ldots, p$; $j = 1, 2$. We will assume that a_i lies in the following set of functions

$$\mathcal{F}_i \;=\; \left\{ f = \sum_k \alpha_{lk}\phi_{lk} + \sum_{j,k} \beta_{jk}\psi_{jk} \;\middle|\; \|\alpha_{l.}\|_\infty \leq C_{i1},\; \|\beta_{..}\|_{m_i,p_i,q_i} \leq C_{i2} \right\},$$

where

$$\|\beta_{..}\|_{m,p,q} \;=\; \left(\sum_{j \geq l} \left[2^{jsp} \sum_{k \in I_j} |\beta_{jk}|^p \right]^{q/p} \right)^{1/q},$$

$s = m + 1/2 - 1/p$.

To have enough regularity, we restrict ourselves to

(A5) $\widetilde{s}_i > 1$ where $\widetilde{s}_i = m_i + 1/2 - 1/\widetilde{p}_i$, with $\widetilde{p}_i = \min\{p_i, 2\}$.

In the case of normally distributed coefficients $\widetilde{\beta}^{(i)}_{jk} \sim N(\beta^{(i)}_{jk}, \sigma^2)$ a very popular method is to apply thresholds $\lambda = \sigma\sqrt{2\log n}$, where n is the number of these coefficients. As shown in Donoho *et al.* (1995), the application of these thresholds leads to a near-optimal estimator in a wide variety of smoothness classes. Because of the heteroscedasticity of the empirical coefficients in our case, we have to modify the above rule slightly. Let $\mathcal{J}_T = \left\{(j,k) \,\middle|\, l \leq j,\; 2^j \leq T^{1/2},\; k \in I_j\right\}$ and let σ^2_{ijk} be the variance of the empirical coefficient $\widetilde{\beta}^{(i)}_{jk}$. Then any threshold λ_{ijk} satisfying

$$\sigma_{ijk}\sqrt{2\log(\#\mathcal{J}_T)} \;\leq\; \lambda_{ijk} \;=\; O(T^{-1/2}\sqrt{\log(T)}) \tag{2.2}$$

would be appropriate. Particular such choices are the "individual thresholds"

$$\lambda_{ijk} \;=\; \sigma_{ijk}\sqrt{2\log(\#\mathcal{J}_T)}$$

and the "universal threshold"

$$\lambda^{(i)}_T \;=\; \sigma^{(i)}_T\sqrt{2\log(\#\mathcal{J}_T)}, \quad \sigma^{(i)}_T \;=\; \max_{(j,k)\in\mathcal{J}_T}\{\sigma_{ijk}\}.$$

Let $\widehat{\lambda}_{ijk}$ be estimators of λ_{ijk} or $\lambda^{(i)}_T$, respectively, which satisfy at least the following minimal condition

(A6) (i) $\sum_{(j,k)\in\mathcal{J}_T} P\left(\widehat{\lambda}_{ijk} < \gamma_T\lambda_{ijk}\right) \;=\; O(T^{\eta})$, where $\eta < 1/(2m_i + 1)$ for some $\gamma_T \to 1$,

(ii) $\sum_{(j,k)\in\mathcal{J}_T} P\left(\widehat{\lambda}_{ijk} > CT^{-1/2}\sqrt{\log(T)}\right) \;=\; O(T^{-1})$.

With such thresholds $\widehat{\lambda}_{ijk}$ we build the estimator

$$\widehat{a}_i(u) \;=\; \sum_{k\in I_l^0} \widetilde{\alpha}_{lk}^{(i)} \phi_{lk}(u) \;+\; \sum_{(j,k)\in\mathcal{J}_T} \delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \widehat{\lambda}_{ijk})\psi_{jk}(u), \tag{2.3}$$

where $\delta^{(\cdot)}$ stands for $\delta^{(h)}$ or $\delta^{(s)}$, respectively.

Finally we like to impose an additional condition on the matrix D being defined by (6.4) in Section 6.1. Basically, this matrix is the analog to the $p \times (T-p)$ matrix $((X_{n-m}))_{n=p+1,\dots,T-p;m=1,\dots,p}$, as arising in the classical Yule-Walker-equations, which describe the corresponding least squares problem for a stationary $AR(p)$-process $\{X_t\}$.

Here, we assume additionally that

(A7) $\mathbb{E}\|(D'D)^{-1}\|^{2+\delta} \;=\; O(T^{-2-\delta})$

for some $\delta > 0$.

Theorem 2.1. *Assume (A1) through (A7). Then*

$$\sup_{a_i\in\mathcal{F}_i} \left\{ \mathbb{E}\|\widehat{a}_i - a_i\|_{L_2[0,1]}^2 \right\} \;=\; O\left((\log(T)/T)^{2m_i/(2m_i+1)} \right).$$

Remark 2. If only (A1) through (A6) are fulfilled, we can still prove that

$$\sup_{a_i\in\mathcal{F}_i} \left\{ \mathbb{E}\left(\|\widehat{a}_i - a_i\|_{L_2[0,1]}^2 \wedge C \right) \right\} \;=\; O\left((\log(T)/T)^{2m_i/(2m_i+1)} \right)$$

holds. Even without (A7) we can show that $D'D$ is close to its expectation $\mathbb{E}D'D$, and hence $\lambda_{\min}(D'D)$ is bounded away from zero, except for an event with very small probability. To take this event into account, the somewhat unusual truncated loss function is introduced.

It is known that the rate $T^{-2m/(2m+1)}$ is minimax for estimating a function with degree of smoothness m in a variety of settings (regression, density estimation, spectral density estimation). Although we do not have a rigorous proof for its optimality in the present context, we conjecture that we cannot do better in estimating the a_i's.

Analogously to Donoho, Johnstone (1992) we can get exactly the rate $T^{-2m_i/(2m_i+1)}$ by the use of level-dependent thresholds $\lambda^{(i)}(j,T,\mathcal{F}_i)$. These thresholds however would depend on the assumed degree of smoothness m_i and it seems to be difficult to determine them in a fully data-driven way. In contrast, the "*log*-thresholds" are much easier to apply, with the small loss of a logarithmic factor in the rate. This simple threshold scheme is possible because it does not aim to achieve the optimal trade-off between bias and variance of the estimator. Rather it is based on a slightly conservative significance test applied to the empirical coefficients.

3. Statistical properties of the empirical coefficients

Before we prove the main theorem in the next section, we give an exact definition of the empirical coefficients and state some statistical properties of them.

First note that our estimator, as a truncated orthogonal series estimator with nonlinearly modified empirical coefficients, involves two smoothing methodologies: one part of the smoothing is due to the truncation above some level j^*. Whereas such a truncation amounts to some linear, spatially not adaptive technique, the more important smoothing is due to the pre-test like thresholding step applied to the coefficients below the level j^*. This step aims at selecting those coefficients which are in absolute value significantly above the noise level and sorting the others out.

From the definition of the Besov norm we obtain that (cf. Theorem 8 in Donoho *et al.* (1995))

$$\sup_{a_i \in \mathcal{F}_i} \left\{ \sum_{j \geq j^*} \sum_k |\widetilde{\beta}_{jk}^{(i)}|^2 \right\} \;=\; O(2^{-2j^* \widetilde{s}_i}), \tag{3.1}$$

where $\widetilde{s}_i = m_i + 1/2 - 1/\min\{p_i, 2\}$. Hence, our loss due to the truncation is of order $T^{-2m_i/(2m_i+1)}$, if j^* is chosen such that $2^{-2j^* \widetilde{s}_i} = O(T^{-2m_i/(2m_i+1)})$. It can be shown by simple algebra that j^* with $2^{j^*-1} \leq T^{1/2} < 2^{j^*}$ is large enough for all smoothness classes from the Besov scale with $\widetilde{s}_i \geq m_i/(m_i + 1/2)$.

We define the empirical coefficients simply as a least squares estimator, i.e. as a minimizer of

$$\sum_{t=p+1}^{T} \left(X_{t,T} \,+\, \sum_{i=1}^{p} \left[\sum_{k \in I_l^0} \alpha_{lk}^{(i)} \phi_{lk}(t/T) \,+\, \sum_{j=l}^{j^*-1} \sum_{k \in I_j} \beta_{jk}^{(i)} \psi_{jk}(t/T) \right] X_{t-i,T} \right)^2 . \tag{3.2}$$

Since $\{\phi_{lk}\}_k \cup \{\psi_{jk}\}_{l \leq j \leq j^*-1,k}$ form a basis of the subspace V_{j^*} of $L_2[0,1]$, this amounts to an approximation of a_i in just this space V_{j^*}.

A first observation about the statistical behavior of the empirical coefficients is stated by the following assertion.

Proposition 3.1. *Assume (A1) through (A7). Then*

(i) $\quad \mathbb{E}(\widetilde{\alpha}_{lk}^{(i)} - \alpha_{lk}^{(i)})^2 \;=\; O(T^{-1}),$

(ii) $\quad \mathbb{E}(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2 \;=\; O(T^{-1})$

hold uniformly in i, k and $j < j^*$.

In view of the nonlinear structure of the estimator, the above assertion will not be strong enough to derive an efficient estimate for the rate of the risk of the estimator. If the empirical coefficients were Gaussian, then the number of $O(2^{j^*})$ coefficients would be dramatically reduced by thresholding with thresholds that are larger by a factor of $\sqrt{2\log(\#\mathcal{J}_T)}$ than the noise level. If we want to tune this thresholding method in accordance to our particular case with non-Gaussian coefficients, we have to investigate the tail behavior of them. Hence, we state asymptotic normality of the coefficients with a special emphasis on moderate and large deviations. To prove the following theorem we decompose the empirical coefficients in a certain quadratic form and some remainder terms of smaller order of magnitude. Then we derive upper estimates for the cumulants of these quadratic forms, which provide asymptotic normality in terms of large deviations due to a lemma by Rudzkis, Saulis and Statulevicius (1978), see Lemma 5.2 in Section 5.

It turns out that we can state asymptotic normality for empirical coefficients $\widetilde{\beta}_{jk}^{(i)}$ with (j,k) from the following set of indices. Let, for arbitrarily small $\delta > 0$,

$$\widetilde{\mathcal{J}}_T \;=\; \left\{(j,k)\, \middle|\, 2^j \geq T^\delta, j < j^*, k \in I_j \right\}.$$

Proposition 3.2. *Assume (A1) through (A6). Then*

$$P((\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})/\sigma_{ijk} \geq x) \;=\; (1-\Phi(x))(1+o(1)) \;+\; O(T^{-\lambda})$$

uniformly in $(j,k) \in \widetilde{\mathcal{J}}_T, x \in \mathbb{R}$ *for arbitrary* $\lambda < \infty$.

We now derive the asymptotic variances of the $\widetilde{\beta}_{jk}^{(i)}$'s. For simplicity of notation we identify $\psi_1, \ldots, \psi_\Delta$ with $\phi_{l1}, \ldots, \phi_{l,2^j+N}, \psi_{l1}, \ldots, \psi_{l,2^l}, \ldots, \psi_{j^*-1,1}, \ldots, \psi_{j^*-1,2^{j^*-1}}$ and $\widetilde{\theta}_1^{(i)}, \ldots, \widetilde{\theta}_\Delta^{(i)}$ with $\widetilde{\alpha}_{l1}^{(i)}, \ldots, \widetilde{\alpha}_{l,2^l+N}^{(i)}, \ldots, \widetilde{\beta}_{l1}^{(i)}, \ldots, \widetilde{\beta}_{l,2^l}^{(i)}, \ldots, \widetilde{\beta}_{j^*-1,1}^{(i)}, \ldots, \widetilde{\beta}_{j^*-1,2^{j^*-1}}^{(i)}$, respectively.

Furthermore, let

$$c(s,k) := \int_{-\pi}^{\pi} \frac{\sigma^2(s)}{2\pi} \left| 1 \;+\; \sum_{j=1}^{p} a_j(s) \exp(i\lambda j) \right|^{-2} \exp(i\lambda k)\, d\lambda. \tag{3.3}$$

$c(s,k)$ is the local covariance of lag k at time $s \in [0,1]$ (cf. Lemma 7.1).

Proposition 3.3. *Assume (A1) through (A5) and (A7). Then*

$$var(\widetilde{\theta}_u^{(i)}) \;=\; T^{-1} \left(A^{-1} B A^{-1}\right)_{p(u-1)+i, p(u-1)+i} \;+\; o(T^{-1}), \tag{3.4}$$

where

$$\begin{aligned} A_{p(u-1)+k, p(v-1)+l} \;&=\; \int \psi_u(s)\psi_v(s) c(s, k-l)\, ds, \\ B_{p(u-1)+k, p(v-1)+l} \;&=\; \int \psi_u(s)\psi_v(s) \sigma^2(s) c(s, k-l)\, ds\;. \end{aligned}$$

Furthermore, $A^{-1} B A^{-1} \geq E^{-1}$ *, where*

$$E_{p(u-1)+k, p(v-1)+l} \;=\; \int \psi_u(s)\psi_v(s) (\sigma^2(s))^{-1} c(s, k-l)\, ds.$$

The eigenvalues of E *are uniformly bounded.*

Remark 3. The above form of A and B suggests different estimates for the variances of $\widetilde{\theta}_u^{(i)}$ and therefore also for the thresholds. One possibility is to use (3.4) and plug in a preliminary estimate ($\sigma^2(s)$ may be estimated by a local sum of squared residuals). Another possibility is to use a nonparametric estimate of the local covariances $c(s,k)$. However, these suggestions require more investigations.

4. Proof of the main theorem

To simplify the treatment of some particular remainder terms which occasionally arise in the following proofs, as e.g. in the decomposition (6.5), we introduce the following notation.

Definition 4.1. We write

$$Z_t \;=\; \widetilde{O}(\eta_T),$$

if for each $\lambda < \infty$ there exists a $C = C(\lambda)$ such that

$$P\left(|Z_T| > C\eta_T\right) \;\leq\; CT^{-\lambda}.$$

(If we use this notation simultaneously for an increasing number of random variables, we mean the existence of a *universal* constant only depending on λ.)

Proof of Theorem 2.1. Using the monotonicity of $\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},.)$ in the second argument we get

$$\begin{aligned}
&\left(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\widehat{\lambda}_{ijk}) \,-\, \beta_{jk}^{(i)}\right)^2 \\
&\leq\; \begin{cases}
(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2 \,+\, (\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\gamma_T\lambda_{ijk}) \,-\, \beta_{jk}^{(i)})^2, \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \widehat{\lambda}_{ijk} < \gamma_T\lambda_{ijk} \\
(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\gamma_T\lambda_{ijk}) \,-\, \beta_{jk}^{(i)})^2 \,+\, (\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},CT^{-1/2}\sqrt{\log(T)}) \,-\, \beta_{jk}^{(i)})^2, \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \gamma_T\lambda_{ijk} \leq \widehat{\lambda}_{ijk} \leq CT^{-1/2}\sqrt{\log(T)} \\
(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},CT^{-1/2}\sqrt{\log(T)}) \,-\, \beta_{jk}^{(i)})^2 \,+\, (\beta_{jk}^{(i)})^2, \\
\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \widehat{\lambda}_{ijk} > CT^{-1/2}\sqrt{\log(T)}
\end{cases}
\end{aligned}$$

which implies the decomposition

$$
\begin{aligned}
&\mathbb{E}\|\widehat{a_i} - a_i\|^2 \\
\le\ & \sum_k \mathbb{E}(\tilde{\alpha}_{lk}^{(i)} - \alpha_{lk}^{(i)})^2 + \sum_{(j,k)\in\mathcal{J}_T} \mathbb{E}(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \widehat{\lambda}_{ijk}) - \beta_{jk}^{(i)})^2 + \sum_{j\ge j^*}\sum_{k\in I_j}(\beta_{jk}^{(i)})^2 \\
\le\ & \sum_k \mathbb{E}(\tilde{\alpha}_{lk}^{(i)} - \alpha_{lk}^{(i)})^2 \\
& + \sum_{(j,k)\in\mathcal{J}_T} \mathbb{E}(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \\
& + \sum_{(j,k)\in\mathcal{J}_T} \mathbb{E}(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, CT^{-1/2}\sqrt{\log T}) - \beta_{jk}^{(i)})^2 \\
& + \sum_{(j,k)\in\mathcal{J}_T} EI\left(\widehat{\lambda}_{ijk} < \gamma_T\lambda_{ijk}\right)(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2 \\
& + \sum_{(j,k)\in\mathcal{J}_T} (\beta_{jk}^{(i)})^2 P\left(\widehat{\lambda}_{ijk} > CT^{-1/2}\sqrt{\log T}\right) \\
& + \sum_{j\ge j^*}\sum_{k\in I_j}(\beta_{jk}^{(i)})^2 \\
=\ & S_1 + \ldots + S_6.
\end{aligned}
\tag{4.1}
$$

By (i) of Proposition 3.1 we get immediately

$$
S_1 = O(T^{-1}). \tag{4.2}
$$

Let $(j,k)\in\widetilde{\mathcal{J}}_T$. We choose a constant γ_{ijk} such that

$$
\begin{aligned}
\delta^{(\cdot)}(\beta, \gamma_T\lambda_{ijk}) &\ge \beta_{jk}^{(i)}, \quad \text{if} \quad \beta - \beta_{jk}^{(i)} > \gamma_{ijk}, \\
\delta^{(\cdot)}(\beta, \gamma_T\lambda_{ijk}) &\le \beta_{jk}^{(i)}, \quad \text{if} \quad \beta - \beta_{jk}^{(i)} < \gamma_{ijk}.
\end{aligned}
$$

(W.l.o.g. we assume $\delta^{(\cdot)}(\gamma_{ijk}, \gamma_T\lambda_{ijk}) \ge \beta_{jk}^{(i)}$.)

Let $\eta_T = CT^{-1/2}\sqrt{\log T}$ for some appropriate C. Then we decompose the terms occurring in the sum S_2 as follows:

$$
\begin{aligned}
S_{21}^{jk} &= \mathbb{E}I\left(\gamma_{ijk} \le \widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)} < \eta_T\right)(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \\
S_{22}^{jk} &= \mathbb{E}I\left(-\eta_T < \widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)} < \gamma_{ijk}\right)(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2
\end{aligned}
$$

and

$$
S_{23}^{jk} = \mathbb{E}I\left(|\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)}| \ge \eta_T\right)(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2.
$$

Using Proposition 3.2 we get, with $\xi_{jk}^{(i)} \sim N(\beta_{jk}^{(i)}, \sigma_{ijk}^2)$, due to integration by parts w.r.t. x

$$\begin{aligned}
S_{21}^{jk} &= -\int \left[I\left(\gamma_{ijk} \le x < \eta_T\right) (\delta^{(.)}(\beta_{jk}^{(i)} + x, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \right] d\{P\left(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)} \ge x\right)\} \\
&= \int \{P\left(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)} \ge x\right)\} d\left[I\left(\gamma_{ijk} \le x < \eta_T\right) (\delta^{(.)}(\beta_{jk}^{(i)} + x, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \right] \\
&\quad + P\left(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)} \ge \gamma_{ijk}\right) (\delta^{(.)}(\beta_{jk}^{(i)} + \gamma_{ijk}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \\
&\le C_T \left\{ \int \{P\left(\xi_{jk}^{(i)} - \beta_{jk}^{(i)} \ge x\right)\} d\left[I\left(\gamma_{ijk} \le x < \eta_T\right) (\delta^{(.)}(\beta_{jk}^{(i)} + x, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \right] \right. \\
&\qquad \left. + P\left(\xi_{jk}^{(i)} - \beta_{jk}^{(i)} \ge \gamma_{ijk}\right) (\delta^{(.)}(\beta_{jk}^{(i)} + \gamma_{ijk}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \right\} \\
&\quad + O(T^{-\lambda}) \\
&= C_T \mathbb{E} I\left(\gamma_{ijk} \le \xi_{jk}^{(i)} - \beta_{jk}^{(i)} < \eta_T\right) (\delta^{(.)}(\xi_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \,+\, O(T^{-\lambda})
\end{aligned}$$

for some $C_T \to 1$. Analogously we get

$$S_{22}^{jk} \,\le\, C_T \mathbb{E} I\left(-\eta_T \le \xi_{jk}^{(i)} - \beta_{jk}^{(i)} < \gamma_{ijk}\right) (\delta^{(.)}(\xi_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \,+\, O(T^{-\lambda}).$$

Finally, we have for any δ_1 with $0 < \delta_1 < \delta$ and δ as in (A7) that

$$S_{23}^{jk} \,\le\, \left(P(|\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)}| \ge \eta_T)\right)^{1-2/(2+\delta_1)} \left(\mathbb{E}|\delta^{(.)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}|^{2+\delta_1}\right)^{2/(2+\delta_1)} \,=\, O(T^{-\lambda}),$$

which implies

$$\mathbb{E}\left(\delta^{(.)}(\widetilde{\beta}_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}\right)^2 \,\le\, C_T \mathbb{E}\left(\delta^{(.)}(\xi_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}\right)^2 \,+\, O(T^{-\lambda}). \tag{4.3}$$

From Lemma 1 in Donoho and Johnstone (1994) we can immediately derive the formula (if $var(\xi_{jk}^{(i)}) = \sigma_{ijk}^2$)

$$E\left(\delta^{(.)}(\xi_{jk}^{(i)}, \lambda) \,-\, \beta_{jk}^{(i)}\right)^2 \,\le\, C\left(\sigma_{ijk}^2 \varphi\left(\frac{\lambda}{\sigma_{ijk}}\right)\left(\frac{\lambda}{\sigma_{ijk}} + 1\right) \,+\, \min\{(\beta_{jk}^{(i)})^2, \lambda^2\}\right), \tag{4.4}$$

where φ denotes the standard normal density. This implies, by Theorem 7 in Donoho *et al.* (1995), that

$$\begin{aligned}
&\sum_{(j,k)\in\widetilde{\mathcal{J}}_T} \mathbb{E}(\delta^{(.)}(\xi_{jk}^{(i)}, \gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)})^2 \\
&\quad = O\left(T^{-1}(\#\widetilde{\mathcal{J}}_T)^{1-\gamma_T^2}\sqrt{\log(T)} \,+\, \sum_{(j,k)\in\widetilde{\mathcal{J}}_T} \min\{(\beta_{jk}^{(i)})^2, (\gamma_T\lambda_{ijk})^2\}\right) \\
&\quad = O\left((\log(T)/T)^{2m_i/(2m_i+1)}\right).
\end{aligned}$$

Therefore, in conjunction with (4.3), we obtain that

$$\sum_{(j,k)\in\widetilde{\mathcal{J}}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}\right)^2 \;=\; O\left((\log(T)/T)^{2m_i/(2m_i+1)}\right). \qquad (4.5)$$

Further we get, because of $|\delta^{(\cdot)}(\beta,\lambda)-\beta| \le \lambda$, that

$$\begin{aligned} &\sum_{(j,k)\in\mathcal{J}_T\setminus\widetilde{\mathcal{J}}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}\right)^2 \\ &\le \sum_{(j,k)\in\mathcal{J}_T\setminus\widetilde{\mathcal{J}}_T} [\; 2\mathbb{E}\left(\widetilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)}\right)^2 \;+\; 2(\gamma_T\lambda_{ijk})^2 \;] \\ &= \#(\mathcal{J}_T \setminus \widetilde{\mathcal{J}}_T) \quad O(T^{-1}\log T). \end{aligned}$$

If we define $\widetilde{\mathcal{J}}_T$ in such a way that $\delta < 1/(2m_i+1)$, we get

$$\sum_{(j,k)\in\mathcal{J}_T\setminus\widetilde{\mathcal{J}}_T} \mathbb{E}\left(\delta^{(\cdot)}(\widetilde{\beta}_{jk}^{(i)},\gamma_T\lambda_{ijk}) - \beta_{jk}^{(i)}\right)^2 \;=\; O\left(T^{-2m_i/(2m_i+1)}\right). \qquad (4.6)$$

By analogous considerations we can show that

$$S_3 \;=\; O\left((\log(T)/T)^{2m_i/(2m_i+1)}\right). \qquad (4.7)$$

From (6.14) and (6.21) we have

$$\widetilde{\beta}_{jk}^{(i)} \;-\; \beta_{jk}^{(i)} \;=\; \tilde{O}\left(T^{-1/2}\sqrt{\log(T)} \;+\; 2^{-j/2}T^{-1/2}\log(T)\right),$$

which implies by (A6)(i) and Lemma 7.2 that

$$\begin{aligned} S_4 \;&=\; O\left(T^{-1}(\log(T))^2\right) \sum_{(j,k)\in\mathcal{J}_T} P\left(\widehat{\lambda}_{ijk} < \gamma_T\lambda_{ijk}\right) \\ &\quad + C\sum_{(j,k)\in\mathcal{J}_T} \left(P\left(|\widetilde{\beta}_{jk}^{(i)} \;-\; \beta_{jk}^{(i)}| \;>\; CT^{-1/2}\log(T)\right)\right)^{2/(2+\delta_1)} \left(\mathbb{E}|\widetilde{\beta}_{jk}^{(i)} \;-\; \beta_{jk}^{(i)}|^{2+\delta_1}\right)^{2/(2+\delta_1)} \\ &=\; O\left(T^{-2m_i/(2m_i+1)}\right). \end{aligned} \qquad (4.8)$$

The relation

$$S_5 \;=\; O\left(T^{-2m_i/(2m_i+1)}\right). \qquad (4.9)$$

is obvious, due to (A6)(ii). Finally, it can be shown by simple algebra that

$$S_6 \;=\; O(2^{-2j^*\widetilde{s}_i}) \;=\; O(T^{-2m_i/(2m_i+1)}), \qquad (4.10)$$

which completes the proof. $\square$

5. A GENERAL LEMMA ON THE CUMULANTS OF QUADRATIC FORMS

In this section we list the basic technical lemmas which are necessary to prove asymptotic normality or to find stochastic estimates for quadratic forms. First, we quote a lemma that provides upper estimates for the cumulants of quadratic forms that satisfy a certain condition on their cumulant sums. This result is a generalization of Lemma 2 in Rudzkis (1978), which was formulated specifically for quadratic forms that occur in periodogram-based kernel estimators of a spectral density. We obtain a slightly improved estimate, which turns out to be important, e.g., for certain quadratic forms with sparse matrices.

We consider the quadratic form

$$\eta_T \;=\; \underline{X}_T' A \underline{X}_T,$$

where

$$\begin{aligned} \underline{X}_T &= (X_1, \ldots, X_T)' \\ A &= ((a_{ij}))_{i,j=1,\ldots,T}, \quad a_{ij} \;=\; a_{ji}. \end{aligned}$$

Further, let

$$\xi_T \;=\; \underline{Y}_T' A \underline{Y}_T,$$

where $\underline{Y}_T = (Y_1, \ldots, Y_T)'$ is a zero mean Gaussian vector with the same covariance matrix as $\underline{X}_T$.

Lemma 5.1. *Assume* $\mathbb{E}X_t = 0$ *and, for some* $\gamma \geq 0$,

$$\sup_{1 \leq t_1 \leq T} \left\{ \sum_{t_2, \ldots, t_k = 1}^{T} |cum(X_{t_1}, \ldots, X_{t_k})| \right\} \;\leq\; C^k (k!)^{1+\gamma} \quad \textit{for all } T \textit{ and } k = 2, 3, \ldots$$

Then, for $n \geq 2$ *,*

$$cum_n(\eta_T) \;=\; cum_n(\xi_T) \;+\; R_n,$$

where

$$\begin{aligned} &\textit{(i)} \quad |cum_n(\xi_T)| \;\leq\; var(\xi_T) 2^{n-2} (n-1)! \left[\lambda_{\max}\left(A\, Cov(\underline{X}_T)\right)\right]^{n-2} \\ &\textit{(ii)} \quad R_n \;\leq\; 2^{n-2} C^{2n} ((2n)!)^{1+\gamma} \max_{s,t}\{|a_{st}|\}\, \widetilde{A}\, \|A\|_\infty^{n-2}, \end{aligned}$$

$$\widetilde{A} \;=\; \sum_s \max_t\{|a_{st}|\}, \quad \|A\|_\infty \;=\; \max_s \left\{ \sum_t |a_{st}| \right\}.$$

The proof of this lemma is given in Neumann (1994).

Using the above lemma we obtain useful estimates for the cumulants, which can be used to derive asymptotic normality. For reader's convenience we quote two basic lemmas on the asymptotic distribution of η_T. The first one, which is due to Rudzkis, Saulis and Statulevicius (1978), states asymptotic normality under a certain relation between variance and the higher order cumulants of η_T. Even if such a favorable relation is not given, we can still get estimates for probabilities of large deviations on the basis of the second lemma, which is due to Bentkus and Rudzkis (1980).

Lemma 5.2. *(Rudzkis, Saulis, Statulevicius (1978))*
Assume for some $\Delta_T \to 0$

$$\left| cum_n\left(\eta_T/\sqrt{var(\eta_T)}\right)\right| \leq \frac{(n!)^{1+\gamma}}{\Delta_T^{n-2}} \quad for \quad n = 3, 4, \ldots$$

Then

$$\frac{P\left(\pm(\eta_T - \mathbb{E}\eta_T)/\sqrt{var(\eta_T)} \geq x\right)}{1 - \Phi(x)} \longrightarrow 1$$

holds uniformly over $0 \leq x \leq \nu_T$*, where* $\nu_T = o(\Delta_T^{1/(3+6\gamma)})$.

Lemma 5.3. *(Bentkus, Rudzkis (1980))*
Let

$$|cum_n(\eta_T)| \leq \left(\frac{n!}{2}\right)^{1+\gamma} \frac{H_T}{\overline{\Delta}_T^{n-2}} \quad for \quad n = 2, 3, \ldots$$

Then, for $x \geq 0$ *,*

$$P(\pm\eta_T \geq x) \leq \exp\left(-\frac{x^2}{2[H_T + (x/\overline{\Delta}_T^{1/(1+2\gamma)})^{(1+2\gamma)/(1+\gamma)}]}\right)$$

$$\leq \begin{cases} \exp\left(-\frac{x^2}{4H_T}\right), & if \quad 0 \leq x \leq (H_T^{1+\gamma}\overline{\Delta}_T)^{1/(1+2\gamma)} \\ \exp\left(-\frac{1}{4}(x\overline{\Delta}_T)^{1/(1+\gamma)}\right), & if \quad x \geq (H_T^{1+\gamma}\overline{\Delta}_T)^{1/(1+2\gamma)} \end{cases}$$

6. Derivation of the asymptotic distribution of the empirical coefficients

6.1. Preparatory considerations. Before we turn directly to the proofs of the Propositions 3.1 through 3.3, we represent the empirical coefficients in a form that allows to recognize easily the nature of every remainder term. Note that throughout the rest of the paper, for notational convenience we now omit the double index in the sequence $\{X_{t,T}\}$, i.e. in the following let $X_t := X_{t,T}$.
Although it is essential for our procedure to have a *multiresolution* basis, i.e. empirical coefficients from different resolution levels, it turns out to be easier to analyze the statistical behavior of such coefficients coming from a single level. Since the empirical coefficients of the multiresolution basis can be obtained as linear combinations of coefficients of an appropriate monoresolution basis, we are able to derive the asymptotic distribution of them.

Since both $\{\phi_{l1}, \ldots, \phi_{l,2^l+N}, \psi_{l1}, \ldots, \psi_{l,2^l}, \ldots, \psi_{j^*-1,1}, \ldots, \psi_{j^*-1,2^{j^*-1}}\}$ and $\{\phi_{j^*1}, \ldots, \phi_{j^*,2^{j^*}+N}\}$ are orthonormal bases of the same space V_{j^*}, the minimization of (3.2) is equivalent to that of

$$\sum_{t=p+1}^{T} \left(X_t + \sum_{i=1}^{p} \left[\sum_{k \in I_{j^*}^0} \alpha_{j^*k}^{(i)} \phi_{j^*k}(t/T)\right] X_{t-i}\right)^2 . \qquad (6.1)$$

The solution $\tilde{\alpha} = (\tilde{\alpha}_{j^*1}^{(1)}, \ldots, \tilde{\alpha}_{j^*1}^{(p)}, \ldots, \tilde{\alpha}_{j^*\Delta}^{(1)}, \ldots, \tilde{\alpha}_{j^*\Delta}^{(p)})'$, $\Delta = \# I_{j^*}^0 = 2^{j^*} + N$, can be written as the least squares estimator

$$\tilde{\alpha} \;=\; (D'D)^{-1} D'Y \tag{6.2}$$

in the linear model

$$Y \;=\; D\alpha \,+\, \gamma, \tag{6.3}$$

where

$$Y \;=\; (X_{p+1}, \ldots, X_T)',$$

$$D \;=\; -\begin{pmatrix} \phi_{j^*1}(\frac{p+1}{T})X_p & \cdots & \phi_{j^*1}(\frac{p+1}{T})X_1 & \cdots & \phi_{j^*\Delta}(\frac{p+1}{T})X_p & \cdots & \phi_{j^*\Delta}(\frac{p+1}{T})X_1 \\ \phi_{j^*1}(\frac{p+2}{T})X_{p+1} & \cdots & \phi_{j^*1}(\frac{p+2}{T})X_2 & \cdots & \phi_{j^*\Delta}(\frac{p+2}{T})X_{p+1} & \cdots & \phi_{j^*\Delta}(\frac{p+2}{T})X_2 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{j^*1}(\frac{T}{T})X_{T-1} & \cdots & \phi_{j^*1}(\frac{T}{T})X_{T-p} & \cdots & \phi_{j^*\Delta}(\frac{T}{T})X_{T-1} & \cdots & \phi_{j^*\Delta}(\frac{T}{T})X_{T-p} \end{pmatrix}, \tag{6.4}$$

$$\alpha \;=\; (\alpha_{j^*1}^{(1)}, \ldots, \alpha_{j^*1}^{(p)}, \ldots, \alpha_{j^*\Delta}^{(1)}, \ldots, \alpha_{j^*\Delta}^{(p)})'$$

and

$$\gamma \;=\; (\gamma_{p+1}, \ldots, \gamma_T)'.$$

The residual term in (6.3) can, for $t = p+1, \ldots, T$, be written as

$$\begin{aligned} \gamma_t \;&=\; X_t \,-\, (D\alpha)_{t-p} \\ &=\; -\sum_{i=1}^{p} a_i(t/T)X_{t-i} \,+\, \varepsilon_t \,+\, \sum_{i=1}^{p}\sum_{k\in I_{j^*}^0} \alpha_{j^*k}^{(i)}\phi_{j^*k}(t/T)X_{t-i} \;=\; \sum_{i=1}^{p} R_i(t/T)X_{t-i} \,+\, \varepsilon_t, \end{aligned}$$

where

$$R_i(u) \;=\; -a_i(u) \,+\, \sum_{k\in I_{j^*}^0} \alpha_{j^*k}^{(i)}\phi_{j^*k}(u) \;=\; -\sum_{j\geq j^*}\sum_{k\in I_j} \beta_{jk}^{(i)}\psi_{jk}(u).$$

Using (6.3) we decompose the right-hand side of (6.2) as

$$\begin{aligned} \tilde{\alpha} \;&=\; (D'D)^{-1}D'D\alpha \,+\, (\mathbb{E}D'D)^{-1}D'e \,+\, \left[(D'D)^{-1} - (\mathbb{E}D'D)^{-1}\right] D'e \,+\, (D'D)^{-1}D'S \\ &=\; \alpha \,+\, T_1 \,+\, T_2 \,+\, T_3, \end{aligned} \tag{6.5}$$

where

$$\begin{aligned} e \;&=\; (\varepsilon_{p+1}, \ldots, \varepsilon_T)', \\ S \;&=\; \left(\sum_{i=1}^{p} R_i(\frac{p+1}{T})X_{p+1-i}, \ldots, \sum_{i=1}^{p} R_i(\frac{T}{T})X_{T-i}\right)'. \end{aligned}$$

Because of the abovementioned relation between the two orthonormal bases of V_{j^*}, there exists an orthonormal $(\Delta \times \Delta)$-matrix Γ with

$$(\phi_{l1}, \ldots, \phi_{l,2^l+N}, \psi_{l1}, \ldots, \psi_{l,2^l}, \ldots, \psi_{j^*-1,1}, \ldots, \psi_{j^*-1,2^{j^*-1}})' \;=\; \Gamma(\phi_{j^*1}, \ldots, \phi_{j^*\Delta})'.$$

This implies

$$(\alpha_{j^*1}^{(i)},\ldots,\alpha_{j^*\Delta}^{(i)})\begin{pmatrix}\phi_{j^*1}\\ \vdots\\ \phi_{j^*\Delta}\end{pmatrix} = (\alpha_{j^*1}^{(i)},\ldots,\alpha_{j^*\Delta}^{(i)})\Gamma'\begin{pmatrix}\phi_{l1}\\ \vdots\\ \phi_{l,2^l+N}\\ \psi_{l1}\\ \vdots\\ \psi_{j^*-1,2^{j^*-1}}\end{pmatrix}.$$

Hence, having the least squares estimator $(\widetilde{\alpha}_{j^*1}^{(i)},\ldots,\widetilde{\alpha}_{j^*\Delta}^{(i)})$ according to the basis $\{\phi_{j^*1},\ldots,\phi_{j^*\Delta}\}$, we obtain the least squares estimator in model (3.2) as

$$\left(\widetilde{\alpha}_{l1}^{(i)},\ldots,\widetilde{\alpha}_{l,2^l+N}^{(i)},\widetilde{\beta}_{l1}^{(i)},\ldots,\widetilde{\beta}_{l,2^l}^{(i)},\ldots,\widetilde{\beta}_{j^*-1,1}^{(i)},\ldots,\widetilde{\beta}_{j^*-1,2^{j^*-1}}^{(i)}\right)' = \Gamma\left(\widetilde{\alpha}_{j^*1}^{(i)},\ldots,\widetilde{\alpha}_{j^*\Delta}^{(i)}\right)'.$$

In other words, every empirical coefficient $\widetilde{\beta}_{jk}^{(i)}$ which is part of the solution to (3.2) can be written as

$$\widetilde{\beta}_{jk}^{(i)} = \Gamma_{ijk}'\widetilde{\alpha}, \tag{6.6}$$

where $\|\Gamma_{ijk}\|_{l_2}=1$. (Analogously, $\widetilde{\alpha}_{lk}^{(i)}=\Gamma_{ik}'\widetilde{\alpha}$.)

6.2. Proofs of the Propositions 3.1, 3.2 and 3.3.

Proof of Proposition 3.1. For notational convenience we write down the proof for empirical coefficients $\widetilde{\beta}_{jk}^{(i)}$ only. The proof for the $\widetilde{\alpha}_{lk}^{(i)}$'s is analogous.

According to (6.5) we have

$$\widetilde{\beta}_{jk}^{(i)} = \beta_{jk}^{(i)} + \Gamma_{ijk}'T_1 + \ldots + \Gamma_{ijk}'T_3. \tag{6.7}$$

From (i) and (iii) of Lemma 7.3 we conclude

$$\begin{aligned}\mathbb{E}(\Gamma_{ijk}'T_1)^2 &= \Gamma_{ijk}'(\mathbb{E}D'D)^{-1}\,Cov(D'e)(\mathbb{E}D'D)^{-1}\Gamma_{ijk}\\ &\le \|\Gamma_{ijk}\|_2^2\|(\mathbb{E}D'D)^{-1}\|_2^2\|\,Cov(D'e)\|_2 = O(T^{-1}).\end{aligned} \tag{6.8}$$

The vector Γ_{ijk} has a length of support of $O(2^{j^*-j})$, which implies

$$\sum_l |(\Gamma_{ijk})_l| \le \|\Gamma_{ijk}\|_2\sqrt{\#\{l \mid (\Gamma_{ijk})_l\neq 0\}} = O(2^{(j^*-j)/2}). \tag{6.9}$$

We have, by Taylor expansion of the matrix $(D'D)^{-1}$, $T_2=T_{21}+T_{22}$, where

$$T_{21} = (\mathbb{E}D'D)^{-1}((\mathbb{E}D'D)-D'D)(\mathbb{E}D'D)^{-1}D'e$$

and

$$\|T_{22}\|_2 = \widetilde{O}\left(\|(\mathbb{E}D'D)^{-1}\|_2^3\|(\mathbb{E}D'D)-D'D\|_2^2\|D'e\|_2\right).$$

Using (i) of Lemma 7.3, (7.8) and (7.9) we get

$$\begin{aligned}\|T_{21}\|_\infty &\le \|(\mathbb{E}D'D)^{-1}\|_\infty^2\|(\mathbb{E}D'D)-D'D\|_\infty\|D'e\|_\infty\\ &= \widetilde{O}\left(2^{j^*/2}T^{-1}\log(T)\right).\end{aligned} \tag{6.10}$$

Since we have enough moment assumptions, we obtain the analogous rate, but without the logarithmic factor, for the second moment of $\Gamma'_{ijk}T_{21}$, i.e.

$$\mathbb{E}(\Gamma'_{ijk}T_{21})^2 \;=\; O\left(2^{j^*-j}2^{j^*}T^{-2}\right). \tag{6.11}$$

Further, we have

$$\Gamma'_{ijk}T_{22} \;=\; \tilde{O}\left(2^{3j^*/2}T^{-3/2}\log(T)\right). \tag{6.12}$$

Using (i) of Lemma 7.3 and (i) of Lemma 7.4 we get

$$\|(D'D)^{-1}\|_2 \;\le\; \|(\mathbb{E}D'D)^{-1}\|_2 + \|(D'D)^{-1}-(\mathbb{E}D'D)^{-1}\|_2 \;=\; O(T^{-1}) + \tilde{O}(2^{j^*/2}T^{-3/2}\sqrt{\log(T)}),$$

which yields, in conjunction with Lemma 7.5, that

$$\begin{aligned}\Gamma'_{ijk}T_3 \;&=\; O\left(\|(D'D)^{-1}\|_2\|D'S\|_2\right)\\ &=\; \tilde{O}\left((2^{-j^*\min\{\tilde{s}_i\}} \;+\; T^{-1/2}2^{-j^*\min\{m_i-1/2-1/(2p_i)\}})\sqrt{\log(T)}\right)\\ &=\; \tilde{O}\left(T^{-1/2-\tau}\right)\end{aligned} \tag{6.13}$$

for some $\tau > 0$. Now we infer from (6.7), (6.8) and (6.11) through (6.13) that

$$\mathbb{E}I(\Omega_0)\left((\tilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2\right) \;=\; O(T^{-1}),$$

where Ω_0 is an appropriate event with $P(\Omega_0) \ge 1 - O(n^{-\lambda})$ for $\lambda < \infty$ chosen sufficiently large. This implies in conjunction with Lemma 7.2, with $0 < \delta_1 < \delta$, that

$$\mathbb{E}I(\Omega_0^c)\left((\tilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)})^2\right) \;\le\; \left(\mathbb{E}|\tilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)}|^{2+\delta_1}\right)^{2/(2+\delta_1)}(P(\Omega_0^c))^{1-2/(2+\delta_1)} \;=\; O(T^{-1}),$$

which finishes the proof. $\square$

Proof of Proposition 3.2. It will turn out that the asymptotic distribution of $\tilde{\beta}_{jk}^{(i)} - \beta_{jk}^{(i)}$ is essentially determined by the behavior of $\Gamma'_{ijk}T_1$. By (6.9), (6.10), (6.12) and (6.13) from the proof of Proposition 3.1 we infer that

$$\Gamma'_{ijk}(T_2 + T_3) \;=\; \tilde{O}\left(2^{-j/2}T^{-1/2}\log(T) + T^{-1/2-\kappa}\right) \tag{6.14}$$

for some $\kappa > 0$.

Now we turn to the derivation of the asymptotic distribution of $\Gamma'_{ijk}T_1$. It is clear that, because of the $MA(\infty)$-representation of the process, $\Gamma'_{ijk}T_1$ can be rewritten as $\sum_{u,v} A_{u,v}\varepsilon_u\varepsilon_v$ for some symmetric matrix $A = A(i,j,k)$. In the following, without writing down the explicit form of this matrix, we derive upper estimates for $\|A\|_\infty$ and $\tilde{A} = \sum_u \max_v\{|A_{u,v}|\}$.

We have

$$\begin{aligned}\Gamma'_{ijk}T_1 &= -\sum_{t=p+1}^{T}\varepsilon_t\sum_{l=1}^{p}X_{t-l}\sum_{u=1}^{\Delta}\phi_{j^*u}(t/T)\sum_v\left((\mathbb{E}D'D)^{-1}\right)_{p(u-1)+l,v}(\Gamma_{ijk})_v \\ &= \sum_{l,s}\left[\sum_t \varepsilon_t\varepsilon_{t-l-s}w_t(l,s)\right], \qquad (6.15)\end{aligned}$$

where

$$w_t(l,s) = \gamma_{t-l}(s)\sum_{u=1}^{\Delta}\phi_{j^*u}(t/T)\sum_v\left((\mathbb{E}D'D)^{-1}\right)_{p(u-1)+l,v}(\Gamma_{ijk})_v.$$

If we write the expression in brackets on the right-hand side of (6.15) as $\sum_{ij}\widetilde{W}_{ij}\varepsilon_i\varepsilon_j$, we obtain, by $\sup_v\{|(\Gamma_{ijk})_v|\} = O(2^{-(j^*-j)/2})$

$$\|\widetilde{W}\|_\infty = O\left(T^{-1}\sup_t\{|\gamma_{t-l}(s)|\}2^{j/2}\right). \qquad (6.16)$$

We can also rewrite $w_t(l,s)$ as

$$w_t(l,s) = -\gamma_{t-l}(s)\sum_v(\Gamma_{ijk})_v\sum_u\left((\mathbb{E}D'D)^{-1}\right)_{v,p(u-1)+l}\phi_{j^*u}(t/T),$$

which implies, by $\sum_v|(\Gamma_{ijk})_v| = O(2^{(j^*-j)/2})$ and by $\sum_t\phi_{j^*u}(t/T) = O(2^{-j^*/2}T)$, that

$$\sum_i\sup_j\{|\widetilde{W}_{ij}|\} = \sum_t|w_t(l,s)| = O(2^{-j/2}). \qquad (6.17)$$

Because of (A3), the summation over s does not affect the rates in (6.16) and (6.17), and so does not the (finite) sum over l. Hence, with the notation of Lemma 5.1, we obtain

$$\|A\|_\infty = O(T^{-1}2^{j/2}), \qquad (6.18)$$

$$\tilde{A} = O(2^{-j/2}). \qquad (6.19)$$

Let $(j,k)\in\widetilde{\mathcal{J}}_T$. Using Lemma 5.1 we obtain

$$\left|cum_n(\Gamma'_{ijk}T_1)\right| \le C^nT^{-1}(n!)^{2+2\gamma}(T^{-1}2^{j/2})^{n-2}, \qquad (6.20)$$

which implies by Lemma 5.2

$$P\left(\pm(\Gamma'_{ijk}T_1)/\sigma_{ijk}\ge x\right) = (1-\Phi(x))(1+o(1)) \qquad (6.21)$$

uniformly in $0\le x\le\kappa_T$, $\kappa_T\asymp T^\nu$ for some $\nu>0$. This relation can obviously be extended to $x\in(-\infty,\kappa_T]$.

Recall that

$$\tilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk} = \Gamma'_{ijk}T_1 + \tilde{O}(T^{-1/2-\kappa}) \qquad (6.22)$$

holds for some $\kappa > 0$. Therefore we have for arbitrary $\lambda < \infty$ that

$$P\left(\pm(\Gamma'_{ijk}T_1)/\sigma_{ijk} - CT^{-\kappa} \geq x\right) - CT^{-\lambda}$$
$$\leq P\left(\pm(\widetilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk})/\sigma_{ijk} \geq x\right) \leq P\left(\pm(\Gamma'_{ijk}T_1)/\sigma_{ijk} + CT^{-\kappa} \geq x\right) + CT^{-\lambda},$$

which implies

$$P\left(\pm(\widetilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk})/\sigma_{ijk} \geq x\right) = [1 - \Phi(x)]\,(1 + o(1)) + O\left(|\Phi(x) - \Phi(x + CT^{-\kappa})|\right) + O\left(|\Phi(x) - \Phi(x - CT^{-\kappa})|\right) + O(T^{-\lambda}). \tag{6.23}$$

Fix any $c > 1$. For $x \leq c$ we have obviously

$$|\Phi(x) - \Phi(x + CT^{-\kappa})| \leq CT^{-\kappa}\phi(0) = o(1 - \Phi(x)). \tag{6.24}$$

For $c < x \leq (2\lambda \log T)^{1/2}$ we obtain by a formula for Mill's ratio (see Johnson and Kotz (1970, vol. 2, p. 278)) that

$$\begin{aligned} |\Phi(x) - \Phi(x + CT^{-\kappa})| &\leq CT^{-\kappa}\phi(x) \\ &\leq CT^{-\kappa}x\left(1 - \frac{1}{x^2}\right)^{-1}(1 - \Phi(x)) \\ &\leq CT^{-\kappa}x\left(1 - \frac{1}{c^2}\right)^{-1}(1 - \Phi(x)) = o(1 - \Phi(x)). \end{aligned} \tag{6.25}$$

The third term on the right-hand side of (6.23) can be treated analogously.

For $x > C(2\lambda \log T)^{1/2}$ we have obviously

$$P\left(\pm(\widetilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk})/\sigma_{ijk} \geq x\right) = O(T^{-\lambda}) = (1 - \Phi(x))(1 + o(1)) + O(T^{-\lambda}), \tag{6.26}$$

which completes the proof. $\square$

Proof of Proposition 3.3. Because of $\mathbb{E}T_1 = 0$ we have

$$Cov(T_1) = \mathbb{E}T_1T_1' = (\mathbb{E}D'D)^{-1}\, Cov(D'e)(\mathbb{E}D'D)^{-1},$$

which implies by (ii) and (iii) of Lemma 7.3 that

$$\|\, Cov(T_1) - F^{-1}GF^{-1}\|_\infty = o(T^{-1}),$$

where

$$F = (\{T \int \phi_{j^*u}(s)\phi_{j^*v}(s)c(s, k - l)\, ds\}_{p(u-1)+k,p(v-1)+l})$$

and

$$G = (\{T \int \phi_{j^*u}(s)\phi_{j^*v}(s)\sigma^2(s)c(s, k - l)\, ds\}_{p(u-1)+k,p(v-1)+l}).$$

This yields

$$\begin{aligned} &\|\, Cov(\Gamma T_1) - \Gamma F^{-1}\Gamma'\Gamma G\Gamma'\Gamma F^{-1}\Gamma'\|_\infty \\ &= \|\, Cov(\Gamma T_1) - A^{-1}BA^{-1}\|_\infty = o(T^{-1}). \end{aligned}$$

Further, due to (6.13) we have

$$\mathbb{E}(\Gamma'_{ijk}(T_2 + T_3))^2 \;=\; o(T^{-1}),$$

which proves the first assertion (3.4).

The matrix $\begin{pmatrix} B & A \\ A & E \end{pmatrix}$ is non-negative definite which leads with Theorem 12.2.21(5) of Graybill (1983) to $A^{-1}BA^{-1} \geq E^{-1}$. Furthermore, we have with $x \in \mathbb{C}^{\Delta p}$

$$\begin{aligned} x^*Ex &= \int_0^1 \int_{-\pi}^{\pi} |A(s,\lambda)|^2 (\sigma^2(s))^{-1} \left| \sum_{u,k} x_{p(u-1)+k} \psi_u(s) \exp(i\lambda k) \right|^2 d\lambda \, ds \\ &\leq C \int_0^1 \int_{-\pi}^{\pi} \left| \sum_{u,k} x_{p(u-1)+k} \psi_u(s) \exp(i\lambda k) \right|^2 d\lambda \, ds \\ &= 2\pi C \|x\|^2, \end{aligned}$$

which implies that the eigenvalues of E are uniformly bounded. $\square$

7. Appendix

In order to preserve a clear presentation of our results, we put some of the technical calculations into this separate section. We assume throughout this section that the assumptions (A1) through (A6) are satisfied.

Let $\Sigma_{t,T} = Cov((X_{t-1,T}, \ldots, X_{t-p,T})')$.

Lemma 7.1. *By (A4), with some constants $C_1, C_2 > 0$,*

(i) $\lambda_{\max}(\Sigma_{t,T}) \leq C_2$ *and* $\lambda_{\min}(\Sigma_{t,T}) \geq C_1 + o(1)$ *, where the $o(1)$ is uniform in t,*

(ii) *there exists some function g, with $g(s) \to 0$ as $s \to 0$, such that*

$$\|\Sigma_{t_1,T} \;-\; \Sigma_{t_2,T}\| \;\leq\; g\left(\frac{t_1 - t_2}{T}\right) \quad \textit{for all } t_1, t_2, T,$$

(iii) $c(s, k-l)$ *is uniformly continuous in s and*

$$\lim_{\substack{T\to\infty \\ t/T \to s}} cov(X_{t-l,T}, X_{t-k,T}) \;=\; c(s, k-l).$$

Proof. Completely analogously to the proof of Theorem 2.3 in Dahlhaus (1995) we can show that $X_{t,T}$ has the representation

$$X_{t,T} \;=\; \int_{-\pi}^{\pi} \exp(i\lambda t) A^0_{t,T}(\lambda) \, d\xi(\lambda)$$

with

$$\sup_{t,\lambda} \left| A^0_{t,T}(\lambda) \;-\; A(t/T, \lambda) \right| \;=\; o(1),$$

where $\xi(\lambda)$ is a process with mean zero and orthonormal increments,

$$A^0_{t,T}(\lambda) \;=\; \frac{1}{\sqrt{2\pi}} \sum_{l=0}^{\infty} \gamma_{t,T}(l) \exp(-i\lambda l)$$

and

$$A(s,\lambda) \;=\; \frac{\sigma(s)}{\sqrt{2\pi}} \left(1 \,+\, \sum_{j=1}^{p} a_j(s) \exp(-i\lambda j) \right)^{-1}.$$

Then

$$cov(X_{t-l,T}, X_{t-k,T}) \;=\; \int_{-\pi}^{\pi} \exp(i\lambda(k-l)) A^0_{t-l,T}(\lambda) A^0_{t-k,T}(-\lambda)\, d\lambda.$$

Since $A(s,\lambda)$ is uniformly continuous in s, this is equal to

$$\int_{-\pi}^{\pi} \exp(i\lambda(k-l)) |A(s,\lambda)|^2\, d\lambda \,+\, o(1) \;=\; c(s, k-l) \,+\, o(1), \quad \text{for } \; t/T \to s,$$

which implies (iii). Analogously, we get (ii). Furthermore, we have for $x = (x_1, \dots, x_p) \in \mathbb{C}^p$

$$\begin{aligned} x^* \Sigma_{t,T} x \;&=\; \int_{-\pi}^{\pi} \left| \sum_{j=1}^{p} x_j \exp(-i\lambda j) A^0_{t-j,T}(\lambda) \right|^2 d\lambda \\ &=\; \int_{-\pi}^{\pi} |A(t/T,\lambda)|^2 \left| \sum_{j=1}^{p} x_j \exp(-i\lambda j) \right|^2 d\lambda \,+\, \|x\|^2\, o(1). \end{aligned}$$

Under (A4) there exist constants with $C_1 \le |A(s,\lambda)| \le C_2$ uniformly in s and λ, which implies (i). $\square$

Lemma 7.2. *Assume additionally (A7) and let* $0 < \delta_1 < \delta$. *Then*

(i) $\quad \mathbb{E}|\tilde{\alpha}^{(i)}_{lk} - \alpha^{(i)}_{lk}|^{2+\delta_1} \;=\; O(1),$

(ii) $\quad \mathbb{E}|\tilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk}|^{2+\delta_1} \;=\; O(1)$

hold uniformly in i, k *and* $j < j^*$.

Proof.
(i) In this part we derive estimates for the moments of $\|D'e\|$ and $\|D'S\|$, which will be used later in this proof.

Using the $MA(\infty)$-representation of $\{X_t\}$ we can write $(D'e)_{p(u-1)+k}$ as a quadratic form $\underline{\varepsilon}' A \underline{\varepsilon}$ for some $A = A(p,k)$, where $\underline{\varepsilon} = (\varepsilon_T, \dots, \varepsilon_1, \varepsilon_0, \varepsilon_{-1}, \dots)'$ is an infinite-dimensional vector according to (A3). Since, however, the proof of Lemma 5.1 does not depend on the dimension of the matrix A, we can apply this lemma also to this infinite-dimensional case.

We obtain, using the notation of Lemma 5.1, that

$$\tilde{A} \;=\; O\left(2^{-j^*/2} T\right),$$

$$\max\{|a_{st}|\} \;\le\; \|A\|_\infty \;=\; O\left(2^{j^*/2}\right),$$

which implies

$$\left| cum_n((D'e)_{p(u-1)+k}) \right| \;\le\; C^n (n!)^{2+2\gamma} T (2^{j^*/2})^{n-2} \qquad \text{for} \quad n \ge 2.$$

Since $\mathbb{E}(D'e)_{p(u-1)+k} = 0$, we get, for even s, that

$$\mathbb{E}\left|(D'e)_{p(u-1)+k}\right|^s \;=\; O\left(\sum_{r=1}^{n} \prod_{i_1,\dots,i_r:\, i_1+\dots+i_r=n,\; i_j\geq 1} |\, cum_{i_j}((D'e)_{p(u-1)+k})|\right) \;\leq\; C(s)T^{s/2}.$$

Now we obtain, with $\Delta = O(2^{j^*})$,

$$\begin{aligned}
\mathbb{E}\|D'e\|^s &= E\left(\sum_{u,k}(D'e)^2_{p(u-1)+k}\right)^{s/2} \\
&\leq (\Delta p)^{s/2-1}\sum_{u,k}\mathbb{E}(D'e)^s_{p(u-1)+k} \\
&= O\left((\Delta p)^{s/2}\max_{u,k}\{\mathbb{E}(D'e)^s_{p(u-1)+k}\}\right) \\
&= O\left(2^{j^* s/2}T^{s/2}\right). \qquad (7.1)
\end{aligned}$$

Now we treat the quantity $\|D'S\|$ in an analogous way. $(D'S)_{p(u-1)+k}$ is a quadratic form in $\underline{X} = (X_1, \dots, X_T)'$ with a matrix A , which satisfies, according to (7.11),

$$\begin{aligned}
\widetilde{A} &= O\left(\sum_t |\phi_{j^*u}(t/T)| \sum_i |R_i(t/T)|\right) \\
&= O\left(\sum_i \sqrt{\sum_t \phi_{j^*u}(t/T)^2}\sqrt{\sum_t R_i(t/T)^2}\right) \\
&= O\left(T(2^{-j^*\min\{\widetilde{s}_i\}} + T^{-1/2}2^{-j^*\min\{m_i-1/2-1/(2p_i)\}})\right) \;=\; O(T^{1/2})
\end{aligned}$$

and, by (7.10),

$$\|A\|_\infty \;=\; O\left(2^{j^*/2}\sum_i \|R_i\|_\infty\right) \;=\; O(2^{j^*/2}).$$

Therefore, we get by Lemma 5.1 that

$$\left|cum_n((D'S)_{p(u-1)+k})\right| \;\leq\; C^n(n!)^{2+2\gamma}T(2^{j^*/2})^{n-2} \qquad \text{for} \quad n\geq 2,$$

which implies, in conjunction with $\mathbb{E}(D'S)_{p(u-1)+k} \;=\; O(\widetilde{A}) \;=\; O(T^{1/2})$, that

$$\mathbb{E}\|D'S\|^s \;=\; O\left(2^{j^* s/2}T^{s/2}\right). \qquad (7.2)$$

(ii) According to (6.5) we have $\tilde{\alpha} - \alpha = (D'D)^{-1}(D'e + D'S)$, which yields that

$$\begin{aligned}
\mathbb{E}|\tilde{\beta}^{(i)}_{jk} - \beta^{(i)}_{jk}|^{2+\delta_1} &= \mathbb{E}|\Gamma'_{ijk}(\tilde{\alpha} - \alpha)|^{2+\delta_1} \\
&\leq \mathbb{E}\left(\|(D'D)^{-1}\|_2 \left(\|D'e\|_2 \,+\, \|D'S\|_2\right)\right)^{2+\delta_1} \\
&\leq \left(\mathbb{E}\|(D'D)^{-1}\|^{2+\delta}\right)^{\frac{2+\delta_1}{2+\delta}} \left(\mathbb{E}(\|D'e\| \,+\, \|D'S\|)^{\frac{(2+\delta_1)(2+\delta)}{\delta-\delta_1}}\right)^{1-\frac{2+\delta_1}{2+\delta}} \\
&= O\left(T^{-(2+\delta_1)}\right) O\left((2^{j^*/2}T^{1/2})^{2+\delta_1}\right) \\
&= O\left((2^{j^*/2}T^{-1/2})^{2+\delta_1}\right) \;=\; O(1).
\end{aligned}$$

□

Lemma 7.3. *Let $j^* = j^*(T) \to \infty$ and $j^* = o(T)$. Then*

(i) $\quad \|(\mathbb{E}D'D)^{-1}\|_\infty \;=\; O(T^{-1}),$

(ii) $\quad \|(\mathbb{E}D'D)^{-1} \;-\; (\{T\int \phi_{j^*u}(s)\phi_{j^*v}(s)c(s,k-l)\,ds\}_{p(u-1)+k,p(v-1)+l})^{-1}\|_\infty \;=\; o(T^{-1}),$

(iii) $\quad \|\,Cov(D'e) \;-\; (\{T\int \phi_{j^*u}(s)\phi_{j^*v}(s)\sigma^2(s)c(s,k-l)\,ds\}_{p(u-1)+k,p(v-1)+l})\|_\infty \;=\; o(T)$

hold uniformly in u, v, k, l.

Proof.
(i) Let $M = T\,Diag[M_1, \ldots, M_\Delta]$, where $M_u = \Sigma_t$ for any t with $t/T \in supp(\phi_{j^*u})$. Because of $M^{-1} = T^{-1}\,Diag[M_1^{-1}, \ldots, M_\Delta^{-1}]$ we get by (i) and (ii) of Lemma 7.1 that

$$\|M^{-1}\|_\infty \;=\; O(T^{-1}). \tag{7.3}$$

Further, we have, by $j^* = j^*(T) \to \infty$ and $j^* = o(T)$, that

$$\begin{aligned}
&(\mathbb{E}D'D \;-\; M)_{p(u-1)+k,p(v-1)+l} \\
&= \sum_{t=p+1}^{T} \phi_{j^*u}(\frac{t}{T})\phi_{j^*v}(\frac{t}{T})\,[(\Sigma_t)_{kl} \;-\; (M_u)_{kl}] \\
&\quad + \left[\sum_{t=p+1}^{T} \phi_{j^*u}(\frac{t}{T})\phi_{j^*v}(\frac{t}{T}) \;-\; T\delta_{uv}\right](M_u)_{kl} \\
&= o(T)
\end{aligned} \tag{7.4}$$

hold uniformly in u, v, k, l. Since ϕ_{j^*u} and ϕ_{j^*v} have disjoint support for $|u-v| \geq C$, we get $(\mathbb{E}D'D)_{kl} = 0$ for $|k-l| \geq Cp$. Therefore we obtain by (7.4)

$$\|\mathbb{E}D'D \;-\; M\|_\infty \;=\; o(T). \tag{7.5}$$

Because of (7.3) and (7.5) there exists a T_0 such that

$$\|M^{-1/2}(\mathbb{E}D'D \;-\; M)M^{-1/2}\| \;\leq\; C < 1 \quad \text{for all} \quad T \geq T_0.$$

Therefore, by the spectral decomposition of $(I + M^{-1/2}(\mathbb{E}D'D - M)M^{-1/2})$ the following inversion formula holds:

$$\begin{aligned}(\mathbb{E}D'D)^{-1} &= \left[M^{1/2}\left(I + M^{-1/2}(\mathbb{E}D'D - M)M^{-1/2}\right)M^{1/2}\right]^{-1} \\ &= M^{-1/2}\left[I + \sum_{s=1}^{\infty}(-1)^s(M^{-1/2}(\mathbb{E}D'D - M)M^{-1/2})^s\right]M^{-1/2}, \qquad (7.6)\end{aligned}$$

which implies (i).

(ii) It can be shown in the same way as (7.4) that

$$\|(\mathbb{E}D'D) - (\{T\int \phi_{j^*u}(s)\phi_{j^*v}(s)c(s,k-l)\,ds\}_{p(u-1)+k,p(v-1)+l})\|_\infty = o(T), \qquad (7.7)$$

which implies analogously to (7.6)

$$\begin{aligned}&\|(\mathbb{E}D'D)^{-1} - (\{T\int \phi_{j^*u}(s)\phi_{j^*v}(s)c(s,k-l)\,ds\}_{p(u-1)+k,p(v-1)+l})^{-1}\|_\infty \\ &= \|(\mathbb{E}D'D)^{-1}\sum_{s=1}^{\infty}(-1)^s\left[(\mathbb{E}D'D - (\{\dots\}))(\mathbb{E}D'D)^{-1}\right]^s\|_\infty = o(T^{-1}).\end{aligned}$$

(iii) Obviously we have

$$\mathbb{E}D'e = 0,$$

which implies

$$\begin{aligned}&cov\left((D'e)_{p(u-1)+k}, (D'e)_{p(v-1)+l}\right) \\ &= \sum_{s,t=p+1}^{T}\phi_{j^*u}(\frac{s}{T})\phi_{j^*v}(\frac{t}{T})\mathbb{E}\varepsilon_s\varepsilon_t X_{s-k}X_{t-l} \\ &= \sum_{s=p+1}^{T}\phi_{j^*u}(\frac{s}{T})\phi_{j^*v}(\frac{s}{T})\mathbb{E}\varepsilon_s^2\,\mathbb{E}X_{s-k}X_{s-l} \\ &= T\int \phi_{j^*u}(s)\phi_{j^*v}(s)\sigma^2(s)c(s,k-l)\,ds + o(T).\end{aligned}$$

The corresponding result in the $\|.\|_\infty$-norm follows from the same reasoning leading to (7.5). □

Lemma 7.4. *It holds that*

$$\textit{(i)}\quad \left\|(D'D)^{-1} - (\mathbb{E}D'D)^{-1}\right\|_\infty = \tilde{O}\left(2^{j^*/2}T^{-3/2}\sqrt{\log(T)}\right)$$

$$\textit{(ii)}\quad \|D'e\|_2^2 = \tilde{O}\left(2^{j^*}T\log(T)\right).$$

Proof.

(i) First, observe that by (A2) and (A3)

$$\begin{aligned}
&\sum_{t_2,\dots,t_k=1}^{T} |cum\,(X_{t_1},\dots,X_{t_k})| \\
&= \sum_{t_2,\dots,t_k=1}^{T} \left| cum\left(\sum_{s_1=-\infty}^{t_1} \gamma_{t_1}(t_1-s_1)\varepsilon_{s_1},\dots, \sum_{s_k=-\infty}^{t_k} \gamma_{t_k}(t_k-s_k)\varepsilon_{s_k} \right) \right| \\
&\leq \sum_{s=-\infty}^{t_1} \sum_{t_2,\dots,t_k=s\vee 1}^{T} |\gamma_{t_1}(t_1-s)|\cdots|\gamma_{t_k}(t_k-s)|\,|\,cum_k(\varepsilon_s)| \\
&\leq \sup_s\{|\,cum_k(\varepsilon_s)|\} \sum_{s=0}^{\infty} |\gamma_{t_1}(s)| \left(\sum_{t=s\vee 1}^{T} |\gamma_t(t-s)| \right)^{k-1} \\
&\leq C^{2k}(k!)^{1+\gamma}.
\end{aligned}$$

We see that

$$(D'D)_{p(u-1)+k,p(v-1)+l} = \sum_{t=p+1}^{T} \phi_{j^*u}(t/T)\phi_{j^*v}(t/T)X_{t-k}X_{t-l}$$

is a quadratic form with a matrix A satisfying, in the notation of Lemma 5.1,

$$\|A\|_\infty = O(2^{j^*}), \quad \widetilde{A} = O(T).$$

This implies by Lemma 5.1 that

$$\begin{aligned}
&\left| cum_n\left((D'D)_{p(u-1)+k,p(v-1)+l} \right) \right| \\
&\leq C^n (n!)^{2+2\gamma}(2^{j^*})^{n-1}T \\
&\leq \left(\frac{n!}{2}\right)^{1+(1+2\gamma)} \frac{H_T}{\overline{\Delta}_T^{n-2}},
\end{aligned}$$

where $H_T \asymp 2^{j^*}T$, $\overline{\Delta}_T \asymp 2^{-j^*}$.

Hence, we get by Lemma 5.3 that

$$P\left(|(D'D)_{p(u-1)+k,p(v-1)+l} - (\mathbb{E}D'D)_{p(u-1)+k,p(v-1)+l}| \geq x \right) \leq \exp\left(-C\frac{x^2}{2^{j^*/2}T} \right)$$

for $0 \leq x \leq (H_T^{1+\gamma}\overline{\Delta}_T)^{1/(1+2\gamma)}$.

Since $(H_T^{1+\gamma}\overline{\Delta}_T)^{1/(1+2\gamma)} \asymp 2^{j^*\gamma/(1+2\gamma)}T^{(1+\gamma)/(1+2\gamma)} \gg 2^{j^*/2}T^{1/2}$, we get

$$(D'D)_{p(u-1)+k,p(v-1)+l} - (\mathbb{E}D'D)_{p(u-1)+k,p(v-1)+l} = \widetilde{O}\left(2^{j^*/2}T^{1/2}\sqrt{\log(T)} \right).$$

Since ϕ_{j^*u} and ϕ_{j^*v} have disjoint support for $|u-v| \geq C$, we immediately obtain

$$\|D'D - \mathbb{E}D'D\|_\infty = \widetilde{O}\left(2^{j^*/2}T^{1/2}\sqrt{\log(T)} \right), \tag{7.8}$$

which yields, in conjunction with (i) of Lemma 7.3,

$$\begin{aligned}
\|(D'D)^{-1} - (\mathbb{E}D'D)^{-1}\|_\infty &\le \|(\mathbb{E}D'D)^{-1}\|_\infty \sum_{s=1}^{\infty} \left(\|D'D - \mathbb{E}D'D\|_\infty \|(\mathbb{E}D'D)^{-1}\|_\infty\right)^s \\
&= O(T^{-1})\tilde{O}\left(2^{j^*/2}T^{1/2}\sqrt{\log(T)}\ T^{-1}\right) \\
&= \tilde{O}\left(2^{j^*/2}T^{-3/2}\sqrt{\log(T)}\right).
\end{aligned}$$

(ii) From similar arguments we obtain

$$(D'e)_{p(u-1)+k} = \tilde{O}\left(T^{1/2}\sqrt{\log(T)}\right), \tag{7.9}$$

which implies (ii). □

Lemma 7.5. *It holds*

$$\|D'S\|_2^2 = \tilde{O}\left(T^2(2^{-2j^*\min\{\tilde{s}_i\}} + T^{-1}2^{-j^*\min\{2m_i-1-1/p_i\}})\log(T)\right).$$

Proof. Because of our assumption $m_i + 1/2 - 1/\tilde{p}_i > 1$ we get

$$\begin{aligned}
\|R_i\|_\infty &= O\left(\sum_{j\ge j^*} 2^{j/2}\max_k\{|\beta_{jk}^{(i)}|\}\right) \\
&= O\left(\sum_{j\ge j^*} 2^{j/2}2^{-js_i}\right) = O\left(2^{-j^*(m_i-1/p_i)}\right)
\end{aligned} \tag{7.10}$$

and

$$\begin{aligned}
TV(R_i) &= O\left(\sum_{j\ge j^*} 2^{j/2}\sum_k |\beta_{jk}^{(i)}|\right) \\
&= O\left(\sum_{j\ge j^*} 2^{j/2}\left(\sum_k |\beta_{jk}^{(i)}|^{p_i}\right)^{1/p_i} 2^{j(1-1/p_i)}\right) \\
&= O\left(\sum_{j\ge j^*} 2^{j/2}2^{-js_i}2^{j(1-1/p_i)}\right) = O\left(2^{-j^*(m_i-1)}\right),
\end{aligned}$$

which implies

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}(R_i(t/T))^2 \;-\; \|R_i\|^2_{L_2[0,1]} \\
&\le \;\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}|R_i(t/T)+R_i(u)|\,|R_i(t/T)-R_i(u)|\;du \\
&= \;\sum_{t} O\left(T^{-1}\|R_i\|_\infty\, TV(R_i)|_{[\frac{t-1}{T},\frac{t}{T})}\right) \\
&= \;O\left(T^{-1}2^{-j^*(2m_i-1-1/p_i)}\right).
\end{aligned}$$

Since we know from Theorem 8 in Donoho *et al.* (1995) that

$$\|R_i\|^2_{L_2[0,1]} \;=\; \sum_{j\ge j^*}\sum_k |\beta^{(i)}_{jk}|^2 \;=\; O\left(2^{-2j^*\widetilde{s}_i}\right),$$

we have that

$$T^{-1}\sum_{t=1}^{T}(R_i(t/T))^2 \;=\; O\left(2^{-2j^*\widetilde{s}_i} \;+\; T^{-1}2^{-j^*(2m_i-1-1/p_i)}\right). \tag{7.11}$$

Now,

$$\begin{aligned}
(D'S)_{p(u-1)+k} \;&=\; \sum_{t=p+1}^{T}\phi_{j^*u}(t/T)X_{t-k}\sum_{i=1}^{p}X_{t-i}R_i(t/T) \\
&=\; \tilde{O}\left(2^{j^*/2}\sqrt{\log(T)}\right)\sum_{t/T\in supp(\phi_{j^*u})}\sum_{i=1}^{p}|R_i(t/T)|,
\end{aligned}$$

which implies

$$\begin{aligned}
\|D'S\|_2^2 \;&=\; \tilde{O}\left(2^{j^*}\log(T)\right)\sum_{i=1}^{p}\sum_{u=1}^{\Delta}\left(\sum_{t/T\in supp(\phi_{j^*u})}|R_i(t/T)|\right)^2 \\
&=\; \tilde{O}\left(2^{j^*}\log(T)\right)\sum_{i=1}^{p}\sum_{u=1}^{\Delta}\left(\sum_{t/T\in supp(\phi_{j^*u})}R_i(t/T)^2\right)T2^{-j^*} \\
&=\; \tilde{O}\left(T^2(2^{-2j^*\min\{\widetilde{s}_i\}} \;+\; T^{-1}2^{-j^*\min\{2m_i-1-1/p_i\}})\log(T)\right).
\end{aligned}$$

□

References

1. Bentkus, R. and Rudzkis, R. (1980). On exponential estimates of the distribution of random variables. *Liet. Mat. Rinkinys* **20**, 15–30. (in Russian)
2. Brillinger, D. R. (1994). Some asymptotics of wavelets in the stationary error case. Technical Report, Department of Statistics, University of California, Berkeley.
3. Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the Interval and Fast Wavelet Transform. *Appl. Comp. Harmonic Anal.* **1**, 54–81.

4. Dahlhaus, R. (1993). Fitting time series models to nonstationary processes. Beiträge zur Statistik 4, Universität Heidelberg.
5. Dahlhaus, R. (1995). On the Kullback-Leibler information divergence of locally stationary processes. Beiträge zur Statistik 27, Universität Heidelberg.
6. Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and statistical estimation. *Appl. Comp. Harmonic Anal.* **1**, 100–115.
7. Donoho, D. L. and Johnstone, I. M. (1992). Minimax estimation via wavelet shrinkage. Technical Report No. 402, Department of Statistics, Stanford University, *Ann. Statist.*, to appear.
8. Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
9. Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion) *J. R. Statist. Soc., Ser. B* **57**, 301–369.
10. Graybill, F. A. (1983). Matrices with Applications in Statistics, 2nd ed. Wadsworth. Belmont, California.
11. Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics. Continuous Univariate Distributions - 1, 2.* Wiley, New York.
12. Johnstone, I. M. and Silverman, B. W. (1994). Wavelet threshold estimators for data with correlated noise. Technical Report, University of Bristol.
13. Meyer, Y. (1991). Ondelettes sur l'intervalle. *Revista Mathematica Ibero-Americana* **7** (2), 115–133.
14. Neumann, M. H. (1994). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, Manuscript.
15. Neumann, M. H. and von Sachs, R. (1995a). Wavelet thresholding: beyond the Gaussian i.i.d. situation. Preprint "Berichte der AG Technomathematik" 133, Universität Kaiserslautern, to appear in: Wavelets and Statistics, Anestis Antoniadis ed., Lectures Notes of Statistics, Springer.
16. Neumann, M. H. and von Sachs, R. (1995b). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Preprint "Berichte der AG Technomathematik" 132, Universität Kaiserslautern
17. Neumann, M. H. and Spokoiny, V. G. (1995). On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods of Statist.* **4**, 137–166.
18. Rudzkis, R. (1978). Large deviations for estimates of spectrum of stationary series. *Lithuanian Math. J.* **18**, 214–226.
19. Rudzkis, R., Saulis, L. and Statulevicius, V. (1978). A general lemma on probabilities of large deviations. *Lithuanian Math. J.* **18**, 226–238.